ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 12 (2025) No. 3, pp. 227-245
DOI:10.33168/JLISS.2025.0314

Comparative Analysis of Machine Learning Algorithms for Academic Performance Prediction Using Oversampling Techniques

Maria Luci Zamora Hernandez[1], Alfredo Daza[2], Ricardo Dario Mendoza Rivera[3], Segundo Nicolas Diestra Sánchez[4], Edgar Gustavo Sparrow Alamo[5], Gonzalo Apaza-Perez[6]

[1]Faculty of Computer and Systems Engineering, Professional School of Computer and Systems Engineering, Universidad Nacional Agraria de la Selva, Tingo Maria, Perú.

[2]Faculty of Engineering, Professional School of Systems Engineering and Informatics, Universidad Nacional del Santa, Nuevo Chimbote, Peru.

[3]Faculty of Engineering, School of Systems Engineering, Universidad Nacional de Trujillo, Trujillo, Peru.

[4] Faculty of Engineering, Professional School of Energy Engineering, Universidad Nacional del Santa, Nuevo Chimbote, Peru.

[5] Faculty of Engineering, Professional School of Civil Engineering, Universidad Nacional del Santa, Nuevo Chimbote, Perú.

[6] Graduate School, Professional School of Systems Engineering, Universidad Nacional del Juliaca, UNAJ, Juliaca, Peru

adaza@uns.edu.pe (Corresponding author)

Abstract. Predicting academic performance has become increasingly critical for educational institutions seeking to implement proactive student support strategies and improve retention rates through early intervention programs. While machine learning approaches demonstrate considerable promise for identifying at-risk students, the impact of class imbalance correction methods and algorithm selection on prediction accuracy remains inadequately understood, particularly within Latin American higher education contexts where institutional and cultural factors may influence the applicability of predictive models developed in other regions. This study systematically evaluates the effectiveness of six machine learning algorithms—Decision Tree, Support Vector Machine, Bayesian Networks, K-Nearest Neighbors, Logistic Regression, and XGBoost—combined with oversampling techniques to predict multi-class academic performance in a Peruvian public university. From an initial dataset of 4,584 student records and 26 attributes, data cleaning reduced it to 2,656 records and 11 features. Class imbalance was addressed using oversampling to ensure equal representation across the academic performance categories. Among the tested algorithms, XGBoost achieved the highest performance with 98.10% accuracy, 98.09% precision, 98.08% recall, and a 98.08% F1-score. These results suggest that ensemble methods, particularly XGBoost, offer superior predictive capability when paired with data balancing techniques. This study is limited to a single-institution dataset; future research should validate the findings using multi-institutional data.

Keywords: machine learning, prediction, academic performance, oversampling.

1. Introduction

The objective of every academic institution is to provide quality teaching to its students, and while one of the necessary indicators of education is academic performance, this is one of the highest priorities for any study center (Bithari et al., 2020). However, a growing trend of declining student performance is emerging, raising serious concerns for education systems (Gutiérrez-de-Rozas et al., 2022). In fact, a study conducted by the Organisation for Economic Co-operation and Development (2019) revealed that 71.8% of students in the Philippines were low-achieving in three core subjects, compared to just 1.1% of students in Shanghai, Zhejiang, Jiangsu, and Beijing.

Academic performance is influenced by multiple factors, including students' learning styles, teacher quality, family background, infrastructure, and peer effects (Briones et al., 2022). Improving academic performance is a major challenge for universities, especially in increasingly competitive environments where outcomes impact reputation and student retention (Ha et al., 2020). Performance is shaped by academic, socio-economic, and institutional factors, requiring early identification of at-risk students to enable timely support. Predictive analytics based on student data allows for targeted interventions (Soyoye et al., 2023). In low- and middle-income countries (LMICs), student attrition is a persistent issue, particularly among disadvantaged populations. Despite the promise of machine learning, many models are context-limited and fail to account for institutional realities in LMICs, reducing their practical value (Salas & Caldas, 2024). Research must therefore address not only algorithmic accuracy but also scalability and applicability in real educational settings.

Predicting student performance enables early intervention to address learning gaps and supports diverse learning strategies (Teoh et al., 2022). This helps educators monitor and improve the efficiency of the learning process, where current technologies allow data to be analyzed in a didactic way (Nabil et al., 2022). Machine learning refers to a system's ability to learn from both training and test data, enabling the automated creation of models that help analyze and solve specific problems (Janiesch et al., 2021), where the importance of making use of algorithms has become increasingly popular by replacing traditional methods and reducing prediction errors (Baashar et al., 2022).

Machine learning is an effective strategy by having the ability to adapt to the user's needs. So, in the field of education, by extracting student data and assessing gaps in learning, you can identify areas where teachers are outnumbered and create solutions in real ways, such as practice tests in additions (Jagwani, 2019; Sandra et al., 2021). Thus, several machine learning algorithms have been used, including Logistic Regression, J48, Multilayer Perceptron, Naive Bayes, Support Vector Machine, Random Forest, etc., for predicting students' academic performance based on comparisons to obtain their accuracy and other metrics (Balaji et al., 2021). This study uses a dataset from a national university in Peru and applies oversampling techniques to improve model performance in predicting academic performance. For this, several machine learning algorithms that have been studied are presented, and a comparison of them is made, as well as mention of possible future advances. Beyond comparing algorithm accuracy, it emphasizes practical scalability and applicability in resource-limited settings, addressing gaps in current research.

This article is structured as follows: Section II shows the background and related works, Section III explains the methodology, Section IV presents the results and discussion, and finally Section V shows the conclusions, limitations and future work.

2. Literature Review

2.1. Background

2.1.1. Machine learning

Machine learning is an artificial intelligence method used for the creation and classification of profiles, both supervised and unsupervised. It is also helpful in many sectors, such as education, in predicting that a student will drop out of a course, be admitted to a curriculum, or identify set tasks (Zawacki-Richter et al., 2019). So, in the machine learning procedure, there is data marked as unmarked data (Zhu et al., 2023).

2.1.2. Unsupervised Learning

It is one of the most extensive types of machine learning. In this approach, the model is trained on an unlabeled dataset, which involves tasks such as association, coding, clustering, and anomaly detection. Thus, the application of unsupervised learning encompasses everything from intrusion detection and data recovery to student learning prediction (Dridi, 2021).

2.1.3. Reinforcement Learning

It is a method of machine learning (ML) that, instead of providing input and output data, describes the current state of the system, where a goal is specified in a list of allowed actions, and whose ML model undergoes the procedure of achieving that objective, the same that works using the principle of trial and error (Silver et al., 2018).

2.1.4. Classification Machine Learning Techniques

It is a method used for predicting similarity while considering a categorical target variable, making it an essential approach for handling various types of data (Reddy & Babu, 2018). It is worth noting that when it comes to the latter, there are several techniques, these being K-nearest neighbors, Bayesian networks, Decision Tree, Logistic regression and Support Vector Machine, applied by different authors to predict academic performance.

For the study, the Jupyter tool and the Python programming language were employed, along with Microsoft Excel for data processing. Additionally, only 11 out of 26 total attributes were selected for analysis. Finally, various Python algorithms were used to predict students' academic performance for comparison purposes (Baig et al., 2023).

2.1.5. Decision Tree (DT)

It is a technique that is utilized for classification problems characterized by dividing the records of the data into established classes through conditions until leaving the nodes of the pure sheets (1 single class) that help define decision-making (Lee et al., 2022), whose types of nodes are divided as follows: decision nodes and leaf nodes (Kaul et al., 2022). For this analysis, the following formula is used to calculate entropy:

$$Entropy(S) = \sum_{i=1}^{c} \quad - Pi\ log_2 Pi, \tag{1}$$

$$Gain\ (S, A) = Entropy\ (S) - \sum_{v \in\ Values(A)} \frac{|Sv|}{|S|} Entropy\ (Sv) \tag{2}$$

2.1.6. Support Vector Machines (SVM)

It is considered a classification algorithm that combines machine learning techniques, including Random Forest, to create robust models used mostly in classification problems (Sharma et al., 2023; Thanh Noi & Kappas, 2017). The mathematical representation is as follows:

Si $Y_i = +1;\ wx_i + b \geq 1$ (3)

Si $Y_i = -1;\ wx_i + b \leq 1$ (4)

For all *i*; $Y_i(wx_i + b) \geq 1$ (5)

In the equation presented above, x is a vector point and w is the weight. Thus, the data must meet the following criteria: (3) it must always be greater than zero, and (4) the data must be less than zero.

2.1.7. Bayesian networks (BN)

It is a graphical representation of probabilities based on conditions, which takes into account random characteristics and conditional dependencies using graphs, where variables are depicted as nodes (Senekane, 2019).

$$P\left(\frac{X}{Y}\right) = \frac{P\left(\frac{Y}{X}\right) x\, P(X)}{P(Y)} \quad (6)$$

2.1.8. K-nearest neighbors (KNN)

This is a supervised algorithm that makes use of proximity to make classifications on a set of data where the destination is known but not the way to reach it. It should be noted that within this algorithm is the value of k (neighbors to be verified), this value being the most important to predict the efficiency of the technique (Turabieh et al., 2021), represented mathematically as follows:

$$d\,(x, y) = \sqrt{\sum_{i=1}^{n} \quad (x_i - y_i)^2} \quad (7)$$

2.1.9. Logistic regression (LR)

It is a probabilistic model of GLM (extension of linear models), whose probability of success (1) and failure (0) of an event is modeled on the independent variables and through the use of the logit function. This model describes the type of binary response based on the characteristics of the variables (Asanya et al., 2023). In LR, the dependent variable is a binary variable that contains coded data. Mathematically, the logistic regression function is plotted as follows:

$$log\,\frac{p\,(y=1)}{1-(p=1)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k; \quad (8)$$

Where k=1, 2…, n

2.1.10. XGboost

It is a method of machine learning that constantly adds new decision trees for the adjustment of a value with several residual iterations, providing a gradient reinforcement framework, characterized by being applied in learning programs such as Kaggle and Python (Chang et al., 2018; Rusdah & Murfi, 2020). It should be noted that this algorithm seeks to minimize the objective function, and for this the loss function and regularization are combined, as shown below:

$$L\emptyset = \sum_i \quad l\,(\hat{y_i}, y_i) + \sum_k \quad \Omega\,(f_k) \quad (9)$$

2.2. Related Works

El-Keiey et al. (2022), in their study, analyzed the performance of students in relation to the time they spend on their activities, in which they used 3 machine learning algorithms: KNN, Decision Tree, and Random Forest. Taking into account a dataset of 500 students, the data was divided: 70% for training and 30% for testing, where a precision of 85% was obtained, as well as an F1 Score of 84% when using Decision Tree. At the same time, in the study of Almarabeh (2017), it was aimed to measure how intelligence and personality traits impact academic performance, where 5 algorithms were used: K-Nearest Network, Support Vector Machine, Decision Tree, Random Forest, and Naive Bayes, obtaining as results that Decision Tree obtained an accuracy and recall of 90% and a precision and F1-score of 89% when making use of 10 attributes of 300 students of the Faculty of Computing and Artificial Intelligence of a university in Cairo.

An analysis of six algorithms (SVM, Decision Tree, Random Forest, KNN, Naïve Bayes, and Logistic Regression) was carried out; they also proposed a multiclass prediction model based on the Synthetic Minority Oversampling Technique (SMOTE), taking into account 1282 course notes. The SVM algorithm obtains an accuracy, precision, recall, and F-measure of 98.90% using SMOTE oversampling, that is, higher values, unlike feature selection (InfoGainAttributeEval), which had an accuracy of 98.4%, precision of 98.10%, recall of 98.4%, and F-score of 97.9% (Bujang et al., 2021). Similarly, in the study by Sudais et al., (2022), it was intended to predict and identify students who could fail the exams using different machine learning techniques, including: Naive Bayes, SVM, Decision Tree and Neural Networks, whose dataset was composed of 839 students from the National

University of Computer and Emerging Sciences, obtaining as results that when comparing different methods (TBN chain and MT chain) with SVM, the first obtained a better score in precision of 62%, recall of 77%, and F1-score of 69%, depending on the number of classes being worked.

In the study of Altabrawee et al. (2019), four machine learning techniques were used with the aim of building a model that helps in the prediction of the performance of 161 students between the years 2015 and 2016 of the Faculty of Humanities of the Al-Muthanna University (MU). An accuracy of 66.52%, a precision of 70.51%, a recall of 64.27%, and an F1-score of 67.21% were obtained using the Bayesian networks, taking data from 161 students of the graduate program of 2 consecutive years. Similarly, Singh and Pal (2020) developed a model combining the Bagging and Boosting algorithms to predict student performance, using the following machine learning techniques: KNN, Naive Bayes, Decision Tree, and ExtraTREE, for which a dataset of 1000 instances and 22 attributes was used. where the results reported a recall of 81.55%, precision of 82.89%, and F1-score of 72.78% using Naive Bayes; however, Bagging achieved the best accuracy value = 91.76%.

Yağcı (2022), in his study, proposed a model based on machine learning algorithms to predict the grades of the final exams of university students, taking as a data source 1854 students enrolled in the subject of Turkish-I Language of a state university in Turkey in the semester 2019-2020, obtaining as results that KNN obtained the best scores of the metrics: accuracy and recall achieved 69.90%, precision 69.10%, and F1-score 69.4%. In addition, Zulfiker et al. (2020) showed that the KNN algorithm achieved an accuracy of 90.38%, precision of 94.00%, recall of 91.00%, and F1-score of 92.00%.

Likewise, El-Hafeez & Omar (2022), used 10 machine learning algorithms (Ridge Ridge, Nearest Centroid, NB multinomial, Random Forest, Logistic Regression, etc.), to build a prediction model, which within the study divided the data into two sections: training data and test data, among the results obtained show that Logistic Regression obtained the highest scores with respect to the metrics, these being: Accuracy and Recall obtained of 71.90%, precision of 72.60% and F1-Score of 71.60%; At the same time as in the study of Rodríguez et al. (2021), it was shown that the same algorithm had the best results, that is an Accuracy of 84.20%, precision of 54.80%, Recall of 36.00% and F1-Score of 43.00%.

Ghorbani and Ghousi (2020), in their study, made a comparison of several resampling techniques (SMOTE-ENN, Random Over Sampler, SMOTE-Tomek, etc.), using 2 different data sets and using machine learning algorithms, including Random Forest, KNN, and XGBoost, among others, where the results show that XGBoost was the one that obtained high values of the metrics, these being accuracy and recall of 69.24%, precision of 64.85%, and an F1 score of 66%, taking into account 2 countries (Iran and Portugal) and unbalanced data. Asselman et al. (2023) obtained an accuracy of 78.75%, precision of 75.12%, recall of 78.75%,, and F1-score of 73.48%.

Other related works, summarized in Table 3, often limit their analysis to a small number of algorithms. In contrast, our study expands upon these efforts by evaluating a greater variety of algorithms, offering a more detailed and comprehensive comparison to better understand the strengths and weaknesses of each approach.

3. Methodology

3.1. Data Source

For this study, a dataset was obtained from a national university in Peru. The original dataset consisted of 4,548 records and 26 variables, including: 'ID', 'Appname', 'Maternal_ID', 'First_Name', 'Gender', 'Department', 'Province', 'District', 'Date_of_Birth', 'Age', 'Marital_Status', 'Student_Code', 'Pronabec_Beneficiary', 'Admission_Mode', 'Type_of_School_of_Origin', 'Professional_School', 'Year_of_Admission', 'Academic_Year', 'Total_Credits_Taken', 'Total_Credits_Failed', 'Total_Credits_Passed', 'Previous_Weighted_GPA', 'Code_Curriculum', 'Final_Weighted_GPA', 'Total_Courses_Enrolled_In', 'Total_Leveling_Courses_Enrolled_In'. These variables included academic, geographic, and socioeconomic information about students. This dataset served as the basis for analysis and the development of predictive models.

3.2. Data Cleaning and Pre-processing

Initially, rows containing missing values in key fields (e.g., marital status, Pronabec beneficiary status) were identified and removed. After this process, the dataset was reduced to 2,677 records and 17 variables. Subsequently, variables with a high percentage of missing values or low predictive value such as Department, Province, District, Professional_School, Year_of_Admission, and Admission_Mode were removed. These were excluded due to their high cardinality, redundancy, and limited contribution to the target variable.

This selection aimed to improve the efficiency of the machine learning models, reduce dataset complexity, and minimize the risk of overfitting. After filtering, 11 attributes (10 predictors and 1 target variable) were retained for modeling, based on data quality, potential predictive value, and relevance to academic performance. These attributes are detailed in Table 1.

Table. 1: Attributes and details of the academic performance dataset

Cod.	Attribute	Acronym	Description
A1	Sex	Sex	Student's gender
A2	Age	Age	Student's age (Years)
A3	Marital status	MS	Student's status as a determinant of legal status
A4	Type of school of origin	TSO	Place of high school from which the student graduated the previous year
A5	Total credits taken	TCT	Total student credits carried
A6	Total number of credits failed	TCF	Total number of credits that the student failed to pass in the university course of study
A7	Total credits passed	TCP	Total number of credits that the student managed to pass within the university career.
A8	Previous weighted average	PWA	Sum of all credits accumulated by the student during the previous cycle.
A9	Total courses enrolled	TCE	Total number of courses allowed to the student upon fulfillment of payment requirements
A10	Total number of remedial courses	TRC	Total number of courses of students who failed to pass their subjects in a regular manner

During cleaning, records with extreme age values (e.g., over 40, 102, or 122 years) were identified. These cases represented less than 1% of the dataset and were considered outliers, as they fall outside the expected range for traditional university students (typically 18–30 years).

From a statistical perspective, these outliers could distort data distribution and reduce the performance of machine learning models, especially those sensitive to variance, such as decision trees or neural networks. Therefore, these records were excluded to homogenize the sample and ensure that models learned patterns representative of the primary student population. Figure 1 illustrates the age distribution before cleaning.

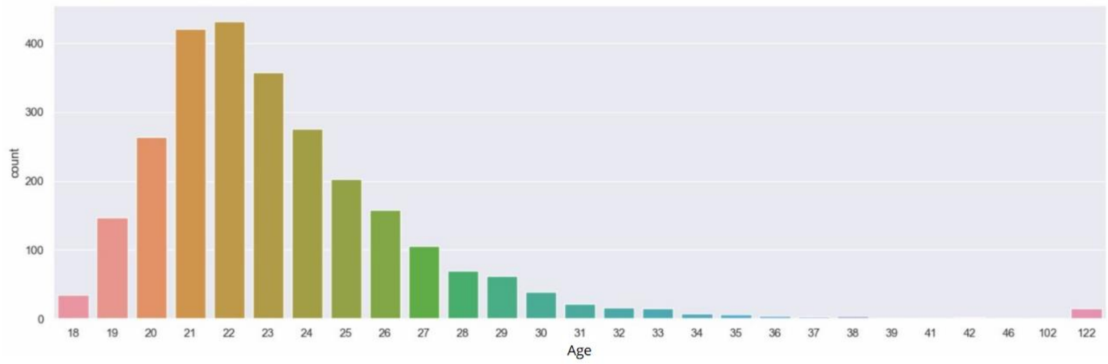

Fig. 1: Age distribution of students before data cleaning

To simplify and standardize inputs for the machine learning models, several transformations were applied. The academic rating variable was discretized into five categories: <11=bad (Malo), >=11 and <13=Regular (Regular), >=13 and <15=Good (Bueno), >=15 and <18=Very good (Muy Bueno), >=18=Excellent (Excelente). Categorical variables were encoded as follows: Gender: F=0 and Marital status: Single=1, Widowed=2, Married=3 and Divorced=4; Type of school: Public=0, Private=1. Figure 2 shows the sequential flow of the proposed methodology.

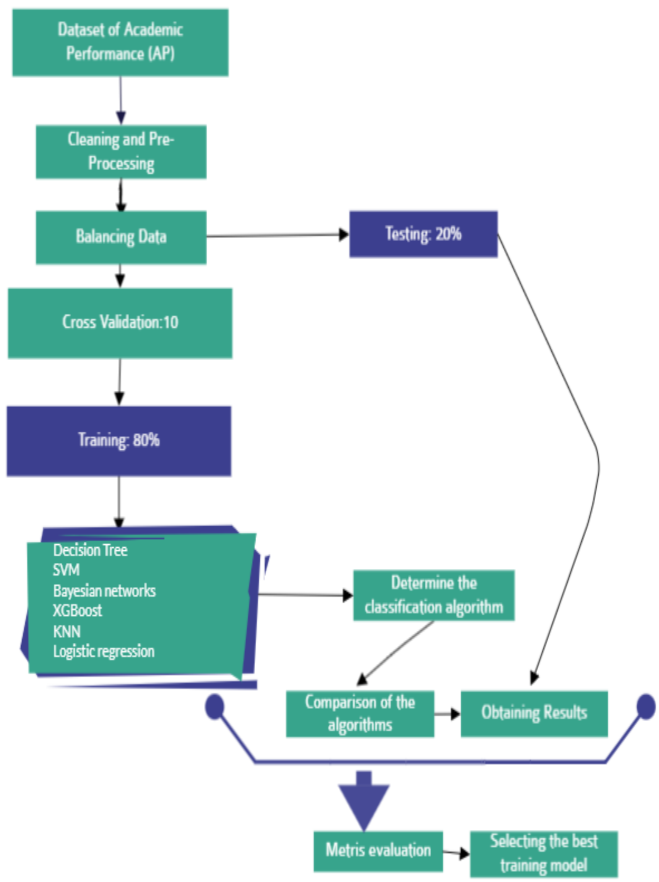

Fig. 2: Sequential chart of proposed model.

3.3. Data Balancing

Initially, the dataset was imbalanced, with class distributions as follows (see Figure 3.a): Good (Bueno) = 623, Bad (Malo) = 677, Very Good (Muy Bueno) = 36, and Regular (Regular) = 1,320. This imbalance can bias machine learning algorithms toward the majority class, reducing predictive accuracy for the minority classes.

To address this issue and improve class representativeness, the oversampling technique was applied. This method increases the number of samples in minority classes by duplicating existing instances. As a result (see Figure 3.b), the dataset was balanced to contain 1,320 instances per class. This adjustment enhanced model learning across all categories and reduced bias toward the majority class. Oversampling was chosen over undersampling to avoid losing valuable information from the larger classes and to maintain the diversity of the dataset.

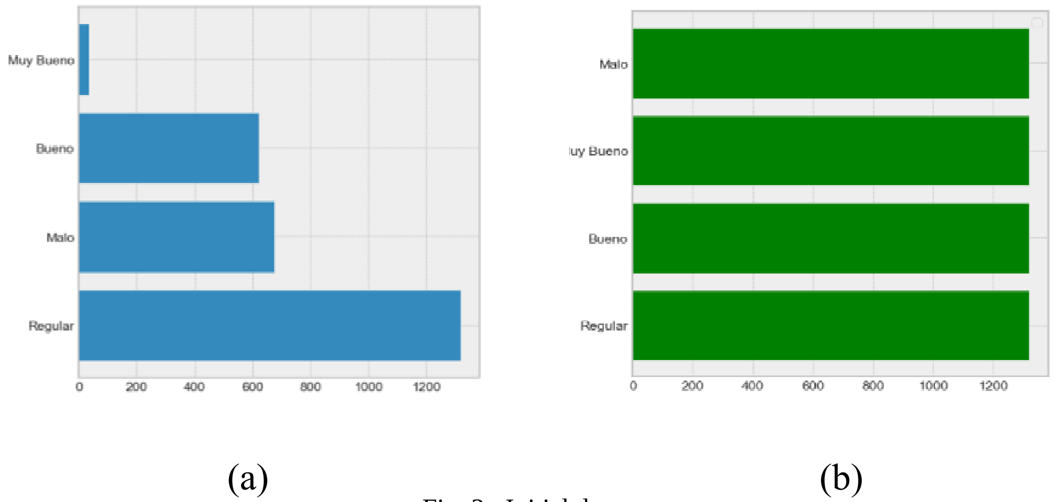

Fig. 3: Initial data

4. Results and Discussion

The purpose of the study was to predict academic performance, taking into account data from a national university in Peru. For the experiment, the data was cleaned and transformed, then the data balancing was performed using the oversampling method, and cross-validation was defined=10, 80% was considered for training and 20% for testing. It was applied to 6 machine learning classification techniques (Decision Tree, SVM, Bayesian networks, KNN, at Logistic regression and XGBoost), obtaining the results of accuracy, precision, recall, and F1-score of each one with respect to the test, all this using the Python tool. The confusion matrix of all the algorithms is presented from Figure 4 to Figure 9, which will allow us to obtain the evaluation metrics.

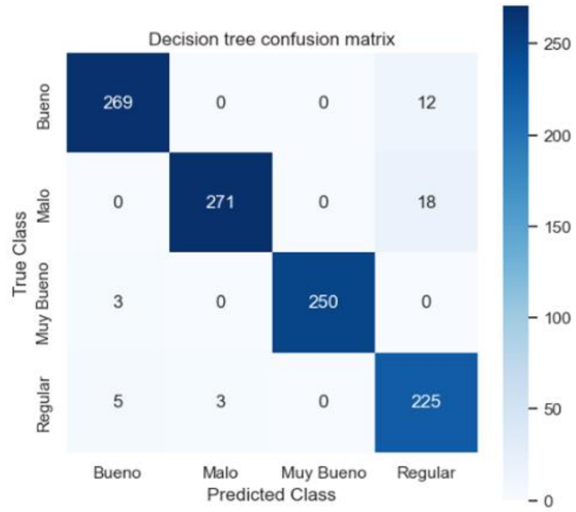

Fig. 4: Decision Tree Confusion Matrix.

Figure 4 illustrates the results for different student performance levels. Among 281 students achieving a “Good (Bueno)” performance, 269 were accurately predicted while 12 were predicted incorrectly. For the 289 students with a “Bad (Malo)” performance, 271 were predicted correctly and 18 incorrectly. In the case of 253 students performing at a “Very good (Muy Bueno)” level, 250 predictions were correct and 3 incorrect. Lastly, out of the 233 students classified as “Regular (Regular),” 225 were predicted accurately while 8 were predicted incorrectly.

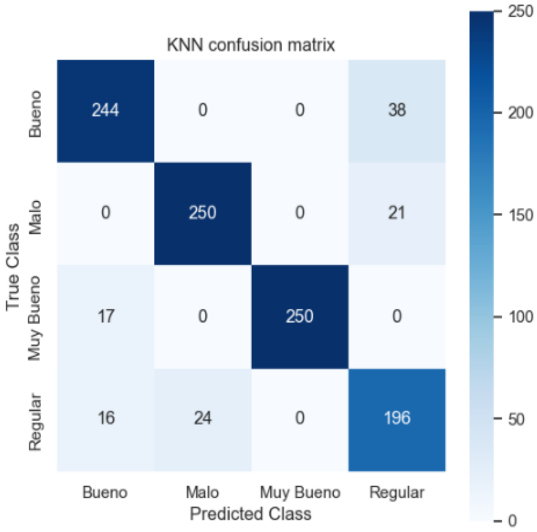

Fig. 5: KNN Confusion Matrix.

Figure 5 illustrates the results for different student performance levels. Among 282 students achieving a “Good (Bueno)” performance, 244 were accurately predicted while 38 were predicted incorrectly. For the 271 students with a “Bad (Malo)” performance, 250 were predicted correctly and 21 incorrectly. In the case of 267 students performing at a “Very good (Muy Bueno)” level, 250 predictions were correct and 17 were incorrect. Lastly, out of the 236 students classified as “Regular (Regular)”, 196 were predicted accurately, while 40 were predicted incorrectly.

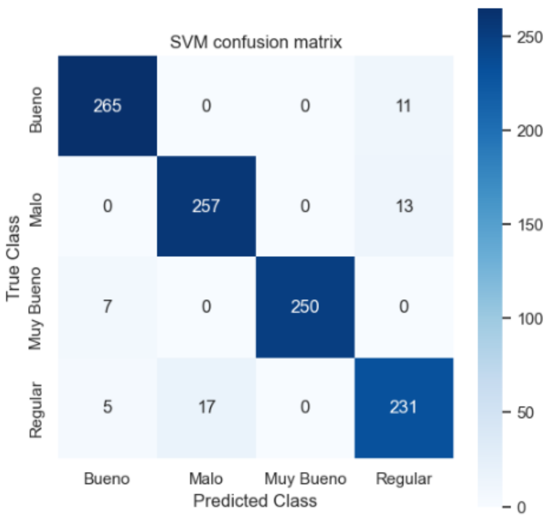

Fig. 6: SVM Confusion Matrix.

Figure 6 indicates the results for different student performance levels. Among 276 students achieving a “Good (Bueno)” performance, 265 were accurately predicted while 11 were predicted incorrectly. For the 270 students with a “Bad (Malo)” performance, 257 were predicted correctly and 13 incorrectly. In the case of 257 students performing at a “Very good (Muy Bueno)” level, 250 predictions were correct and 7 incorrect. Lastly, out of the 253 students classified as "Regular (Regular)" 231 were predicted accurately, while 22 were predicted incorrectly.

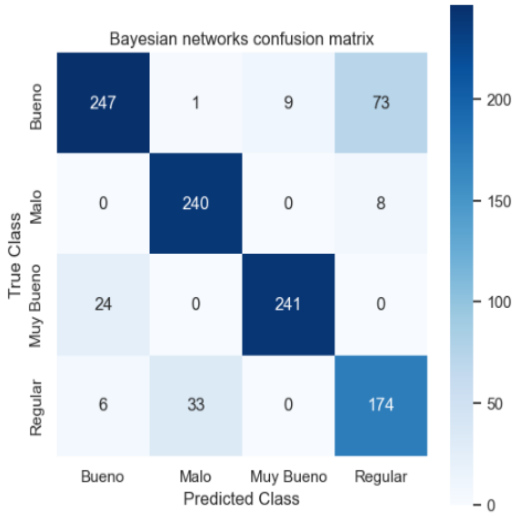

Fig. 7: Bayesian networks confusion matrix.

Figure 7 details the results for different student performance levels. Among 330 students achieving a "Good (Bueno)" performance, 247 were predicted accurately, while 83 were predicted incorrectly. For the 248 students with a "Bad (Malo)" performance, 240 were predicted correctly and 8 incorrectly. In the case of 265 students performing at a "Very good (Muy Bueno)" level, 241 predictions were correct and 24 incorrect. Lastly, out of the 213 students classified as "Regular (Regular)", 174 were predicted accurately while 39 were predicted incorrectly.

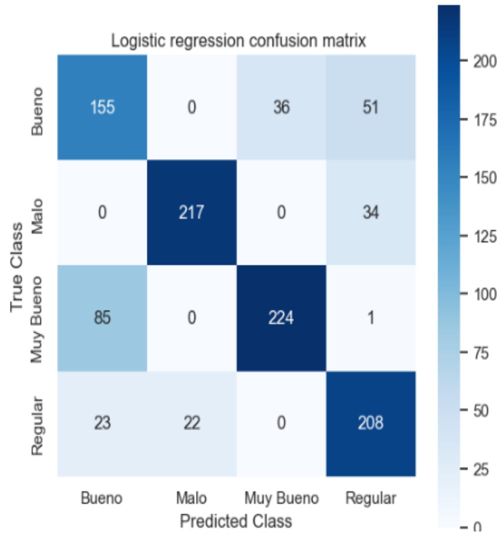

Fig. 8: Logistic regression Confusion matrix.

Figure 8 presents the results for different student performance levels. Among 242 students achieving a "Good (Bueno)" performance, 155 were predicted accurately, while 87 were predicted incorrectly. For the 251 students with a "Bad (Malo)" performance, 217 were predicted correctly and 34 incorrectly. In the case of 310 students performing at a "Very good (Muy Bueno)" level, 224 predictions were correct and 86 incorrect. Lastly, out of the 253 students classified as "Regular (Regular)", 208 were predicted accurately while 45 were predicted incorrectly.

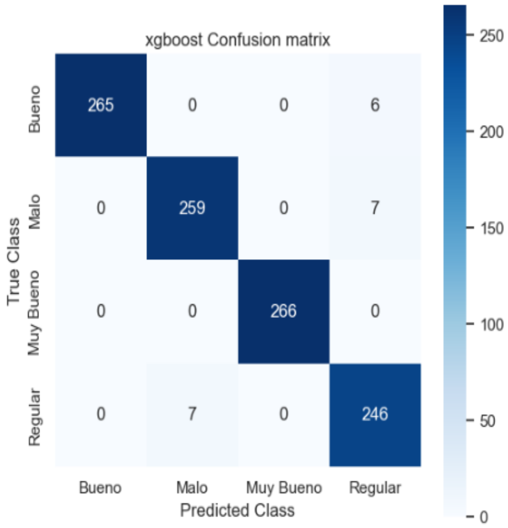

Fig.9: XGBoost Confusion matrix.

Figure 9 displays the results for different student performance levels. Among 271 students achieving a “Good (Bueno)” performance, 265 were accurately predicted while 6 were predicted incorrectly. For the 266 students with a “Bad (Malo)” performance, 259 were predicted correctly and 7 incorrectly. In the case of 266 students performing at a “Very good (Muy Bueno)” level, 266 predictions were correct and 0 incorrect. Lastly, out of the 253 students classified as “Regular (Regular)”, 246 were predicted accurately, while 7 were predicted incorrectly.

4.1. Accuracy (ACC)

This metric represents the number of samples that were correctly classified in relation to the total number of test samples (Srividya et al., 2018). We can express this metric mathematically as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{10}$$

Table 3 and Figure 10 indicate the comparison of the “Accuracy” metric between the different techniques to predict academic performance, which uses 2,656 instances. The graph indicates that the accuracy in the case of the XGBoost algorithm was higher, being 98.10% compared to the other models, where MLP obtained an accuracy of 92.30% (Aggarwal et al., 2019), while KNN obtained a value of 94.07% and ANN of 95.38%; however, it was surpassed by the hybrid model with 98.80% (Deepika & Sathyanarayana, 2020).

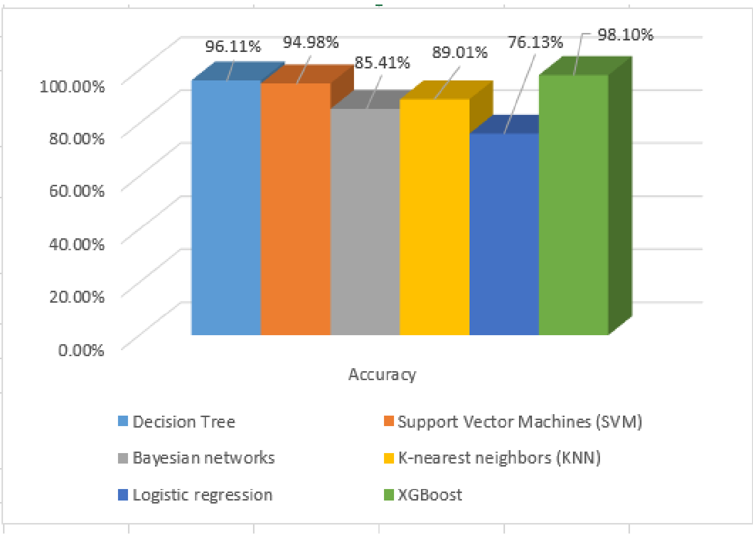

Fig. 10: Accuracy Metric.

4.2. Recall

Recall is the ability of the test to correctly predict students with anxiety. i.e., it is the part of cases predicted as positive (Erickson & Kitamura, 2021). Expressed as follows:

$$Recall = \frac{(TP)}{(TP+FN)}*100 \tag{11}$$

On the other hand, Table 3 and Figure 11 show the results with regard to recall. It was evidenced that XGBoost achieved a high value of 98.08%, unlike Random Forest with 94.90% (Aggarwal et al., 2019), Logistic Regression and KNN with a value of 79.00%, and Naive Bayes with 68.00% (Alturki et al., 2022).

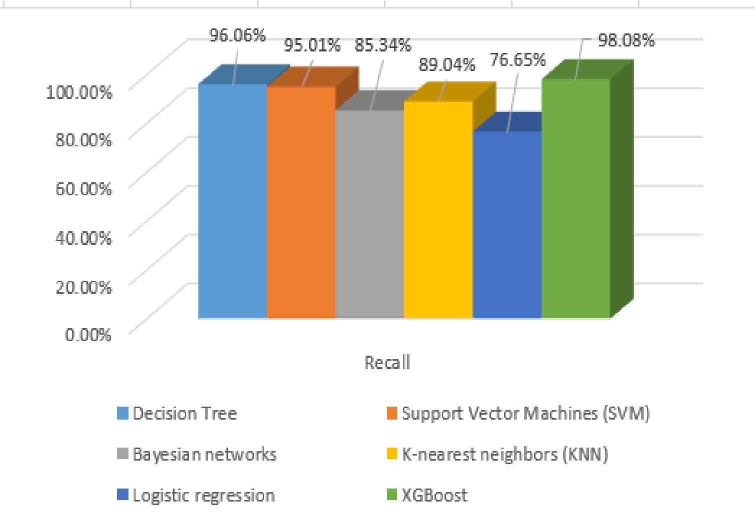

Fig. 11: Recall Metric.

4.3. Precision

It is the metric of a correctly classified class; in other words, it is the relationship between true positives and the number of cases classified as positive (false positives and true positives) (Vakili et al., 2020).

$$Precision = \frac{TP}{TP+FP} \tag{12}$$

Regarding the precision metric, Table 3 and Figure 12 show the results: the XGBoost algorithm obtained a precision of 98.09%, that is, it was the one that obtained the highest value compared to J48 and NNge, which achieved a precision of 95.80% and 92.90%, respectively (Imran et al., 2019). It is also observed that Random Forest achieved a value of 95.00% (Jawad et al., 2022).

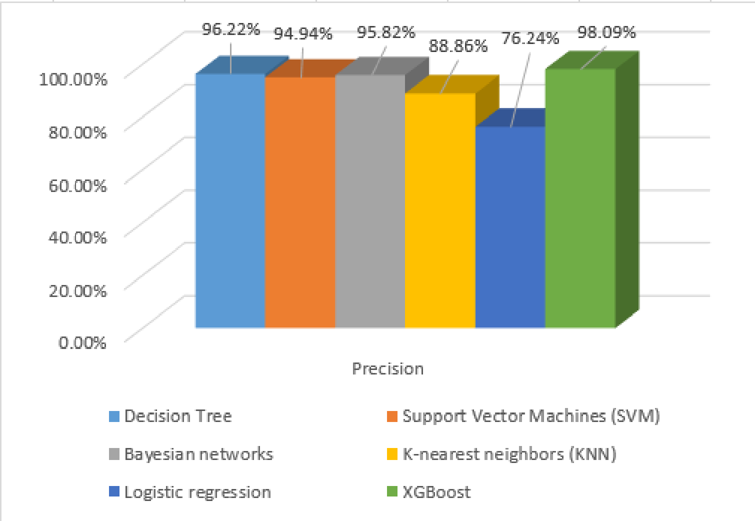

Fig. 12: Precision Metric.

4.4. F1-score

It is a metric that combines the true positive rate (recall) and the positive predictive value (precision), with scores ranging from 0 to 1. This metric provides a balanced measure of accuracy and recall, known as the F1 score (Handelman et al., 2019), represented by the following formula:

$$F1\ Score = \frac{2TP}{(2TP+FP+FN)} \tag{13}$$

Finally, in the proposed model, XGBoost was the one that obtained a high value for the F1-score metric compared to other algorithms, this being 98.08% (see Figure 13). In the case of MLP, a value of 93.40% and NNge of 92.80% (Imran et al., 2019) were obtained; however, they were surpassed by the hybrid model with 99.00% (Deepika & Sathyanarayana, 2020).

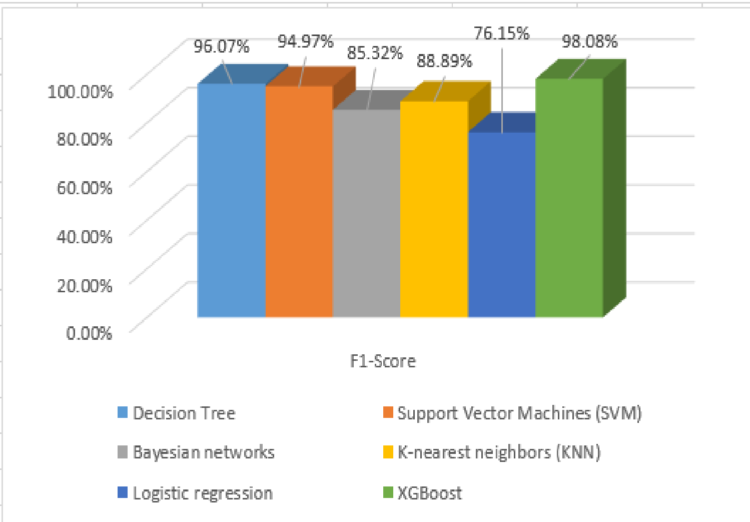

Fig. 13: F1-Score Metric.

Table 2 details the values of the metrics obtained in each technique. The comparative analysis is presented in Table 3 with their respective accuracy, precision, recall, and F1-score scores for the prediction of academic performance, as well as the model proposed by the different authors.

Table. 2: Metrics and corresponding classification technique percentages

	Accuracy	Precision	Recall	F1-Score
Decision Tree	96.11%	96.22%	96.06%	96.07%
Support Vector Machines (SVM)	94.98%	94.94%	95.01%	94.97%
Bayesian networks	85.41%	95.82%	85.34%	85.32%
K-nearest neighbors (KNN)	89.01%	88.86%	89.04%	88.89%
Logistic regression	76.13%	76.24%	76.65%	76.15%
XGBoost	98.10%	98.09%	98.08%	98.08%

The outstanding performance of the XGBoost algorithm, with an accuracy of 98.10%, can be attributed to its ability to handle imbalanced and noisy data through regularization techniques and its boosting-based learning approach, which iteratively corrects classification errors during training. This robustness makes it particularly well-suited for educational contexts, where relationships between variables are often complex and nonlinear. In contrast, models such as logistic regression (76.13%) and KNN (89.01%) showed lower performance, possibly due to their sensitivity to multicollinearity or their reliance on more structured and homogeneous data. These findings suggest that the choice of algorithm is crucial when implementing academic performance prediction systems in educational institutions, as more advanced models like XGBoost may support early identification of at-risk students and enable more effective pedagogical interventions.

Table. 3: Metrics to predict academic performance with different ML techniques

Authors	Technique	Metrics
Imran et al. (2019)	J48	Precision: 95.80%; Recall: 95.80%; F1-Score: 95.80%
	NNge	Precision: 92.90%; Recall: 92.80%; F1-Score: 92.80%
	MLP	Precision: 93.40%; recall: 93.40%; F1-Score: 93.40%
Aggarwal et al. (2019)	Random Forest	Acc: 92.30; Precision: 92.50%; Recall: 94.90%; F1-Score: 93.70%
	MLP	Acc: 92.30; Precision: 95.70%; Recall: 97.40%; F1-Score: 93.80%
Jawad et al. (2022)	Random Forest	Acc: 97.80%; Precision-Recall: 95.00%; F1-Score: 89.80%
Deepika and Sathyanarayana (2020)	KNN	Acc: 93.07%; Precision: 87.00%; Recall: 93.00%; F1-Score: 90.00%
	ANN	Acc: 95.38%; Precision: 96.00%; Recall: 95.00%; F1-Score: 92.00%
	Hybrid model	Acc: 98.80%; Precision: 99.00%; Recall:99.00%; F1-Score: 99.00%
Alturki et al. (2022)	LR	Acc: 80.00%; Precision: 76.00%; Recall: 79.00% ; F1-Score: 77.00%
	RF	Acc: 81.00%; Precision: 75.00%; Recall: 82.00% ; F1-Score: 78.00%
	KNN	Acc: 79.00%; Precision: 74.00%; Recall: 79.00%; F1-Score: 76.00%
	NB	Acc: 78.00%; Precision: 76.00%; Recall: 68.00%; F1-Score: 72.00%

	SVM	Acc: 80.00%; Precision: 78.00%; Recall: 75.00%; F1-Score: 76.00%
	ANN	Acc: 77.00%; Precision: 71.00%; Recall: 74.00%; F1-Score: 73.00%
The proposed model	DT	Acc: 96.11%; Precision: 96.22%; Recall: 96.06%; F1-Score: 96.07%
	SVM	Acc: 94.98%; Precision: 94.94%; Recall: 95.01%; F1-Score: 94.97%
	BN	Acc: 85.41%; Precision: 95.82%; Recall: 85.34%; F1-Score: 85.32%
	KNN	Acc: 89.01%; Precision: 88.86%; Recall: 89.04%; F1-Score: 88.89%
	LR	Acc: 76.13%; Precision: 76.24%; Recall: 76.65%; F1-Score: 76.15%
	XGBoost	Acc: 98.10%; Precision: 98.09%; Recall: 98.08%; F1-Score: 98.08%

Table 3 presents a selection of previous studies that analyze academic performance using machine learning. These works generally focus on a limited set of algorithms. These studies typically evaluate a limited number of algorithms, focusing on the most commonly used classifiers such as Random Forest, SVM, and Neural Networks. The metrics reported—accuracy, precision, recall, and F1-score—provide a comprehensive overview of each model's predictive effectiveness.

In contrast, the present study distinguishes itself by incorporating a wider array of machine learning methods, including traditional models (e.g., Decision Trees, Logistic Regression) and advanced ensemble techniques (e.g., XGBoost). This broader evaluation allows for a more thorough comparison of model performance across diverse algorithmic approaches.

Moreover, by systematically applying the same evaluation metrics across all models, this study ensures consistent benchmarking, which enhances the reliability of the results. This approach not only highlights the strengths and weaknesses of individual algorithms but also supports the identification of the most effective model for academic performance prediction.

The proposed model, as shown in the table, achieves competitive results, particularly with ensemble methods like XGBoost, which outperforms many existing models in terms of accuracy and balanced precision-recall performance. This underlines the importance of exploring a diverse set of algorithms for robust predictive analytics in educational data mining. However, it is important to acknowledge that differences in student populations, institutional contexts, and evaluation protocols may influence the direct comparability of these results. Therefore, such comparisons should be interpreted with caution.

5. Conclusions

The study proposed a methodology and and present a comparative analysis of machine learning techniques using an oversampling approach for predicting academic performance. A total of 11 attributes were considered, and six machine learning techniques were applied: Decision Tree, Support Vector Machine (SVM), Bayesian Networks, K-Nearest Neighbors (KNN), Logistic Regression, and XGBoost. Data preprocessing and balancing were performed prior to model training. Among the techniques tested, XGBoost achieved the highest accuracy, recall, and F1-score.

Given the high accuracy of the XGBoost model and the identification of key influencing variables, it is recommended that educational institutions use these predictions as a complementary tool to design personalized interventions, targeting resources and support to students who exhibit critical indicators such as low prior weighted averages or a high number of failed credits. Additionally, institutions should implement early warning systems to monitor at-risk students and provide timely tutoring or academic advising.

5.1. Limitations and Future Work

Regarding implementation, it is important to consider that while XGBoost offers the highest accuracy, it requires greater computational resources, which might not be feasible for all educational institutions, especially those with limited infrastructure. In such cases, simpler models like Decision Trees or Support Vector Machines provide a reasonable trade-off between accuracy and resource consumption. Institutions should evaluate their available hardware, data volume, and budget before choosing a model. Cloud-based services could be considered to overcome local hardware limitations. Future work could also explore model optimization for scalability and cost reduction.

Additionally, this study did not include a sensitivity or error analysis, nor did it examine model interpretability in depth. In addition, no statistical significance testing was performed to validate whether differences among algorithms are meaningful. Learning curves and confidence intervals were also omitted, which limits the assessment of overfitting and performance reliability. Future research should incorporate these aspects to enhance the transparency, robustness, and fairness of the proposed models, particularly in relation to variables such as school type or socioeconomic background.

It is essential that educators remain involved with human oversight to critically review predictions and avoid negative labeling that could impact student motivation or introduce bias. Furthermore, institutions must uphold strict data privacy and ethical standards, clearly informing students about data usage and preventing adverse consequences from automated performance classifications. By doing so, the predictive model can serve as an ethical and practical resource to enhance educational outcomes and student experiences.

References

Aggarwal, D., Mittal, S., & Bali, V. (2019). Prediction model for classifying students based on performance using machine learning techniques. *International Journal of Recent Technology and Engineering*, *8*(2 Special Issue 7), Article 2 Special Issue 7. https://doi.org/10.35940/ijrte.B1093.0782S719

Almarabeh, H. (2017). Analysis of Students' Performance by Using Different Data Mining Classifiers. *International Journal of Modern Education and Computer Science*, *9*(8), 9-15. https://doi.org/10.5815/ijmecs.2017.08.02

Altabrawee, H., Ali, O. A. J., & Ajmi, S. Q. (2019). Predicting Students' Performance Using Machine Learning Techniques. *Journal of University of Babylon for Pure and Applied Sciences*, *27*(1), 194-205. https://doi.org/10.29196/jubpas.v27i1.2108

Alturki, S., Cohausz, L., & Stuckenschmidt, H. (2022). Predicting Master's students' academic performance: An empirical study in Germany. *Smart Learning Environments*, *9*(1), 38. https://doi.org/10.1186/s40561-022-00220-y

Asanya, K. C., Kharrat, M., Udom, A. U., & Torsen, E. (2023). Robust Bayesian approach to logistic regression modeling in small sample size utilizing a weakly informative student's t prior distribution. *Communications in Statistics - Theory and Methods*, *52*(2), 283-293.

Asselman, A., Khaldi ,Mohamed, & and Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, *31*(6), 3360-3379. https://doi.org/10.1080/10494820.2021.1928235

Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., Hashim, W., & Tiong, S. K. (2022). Toward Predicting Student's Academic Performance Using Artificial Neural Networks (ANNs). *Applied Sciences*, *12*(3), Article 3. https://doi.org/10.3390/app12031289

Baig, M. A., Shaikh, S. A., Khatri, K. K., Shaikh, M. A., Khan, M. Z., & Rauf, M. A. (2023). Prediction of Students Performance Level Using Integrated Approach of ML Algorithms. *International Journal of Emerging Technologies in Learning (iJET)*, *18*(01), Article 01. https://doi.org/10.3991/ijet.v18i01.35339

Balaji, P., Alelyani, S., Qahmash, A., & Mohana, M. (2021). Contributions of Machine Learning Models towards Student Academic Performance Prediction: A Systematic Review. *Applied Sciences*, *11*(21), Article 21. https://doi.org/10.3390/app112110007

Bithari, T. B., Thapa, S., & K.c, H. (2020). Predicting Academic Performance of Engineering Students Using Ensemble Method. *Technical Journal*, *2*(1), Article 1. https://doi.org/10.3126/tj.v2i1.32845

Briones, S. K. F., Dagamac, R. J. R., David, J. D., & Landerio, C. A. B. (2022). Factors Affecting the Students' Scholastic Performance: A Survey Study. *Indonesian Journal of Educational Research and Technology*, *2*(2), Article 2. https://doi.org/10.17509/ijert.v2i2.41394

Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. Md. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. *IEEE Access*, *9*, 95608-95621. https://doi.org/10.1109/ACCESS.2021.3093563

Chang, Y.-C., Chang, K.-H., & Wu, G.-J. (2018). Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing*, *73*, 914-920. https://doi.org/10.1016/j.asoc.2018.09.029

Deepika, K., & Sathyanarayana, N. (2020). Hybrid model for improving student academic performance. *Int J Adv Res Eng Technol*, *11*(10), 768-779.

Dridi, S. (2021). *Supervised Learning—A Systematic Literature Review*. OSF. https://doi.org/10.31219/osf.io/qtmcs

El-Hafeez, T. A., & Omar, A. (2022). *Student Performance Prediction Using Machine Learning Techniques*. https://doi.org/10.21203/rs.3.rs-1455610/v1

El-Keiey, S., ElMenshawy, D., & Hassanein, E. (2022). Student's Performance Prediction based on Personality Traits and Intelligence Quotient using Machine Learning. *International Journal of Advanced Computer Science and Applications*, *13*(9). https://doi.org/10.14569/IJACSA.2022.0130934

Erickson, B. J., & Kitamura, F. (2021). *Magician's Corner: 9. Performance Metrics for Machine Learning Models*. *3*(3). https://doi.org/10.1148/ryai.2021200126

Ghorbani, R., & Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *IEEE Access*, *8*, 67899-67911. https://doi.org/10.1109/ACCESS.2020.2986809

Gutiérrez-de-Rozas, B., López-Martín, E., & Carpintero Molina, E. (2022). Defining the profile of students with low academic achievement: A cross-country analysis through PISA 2018 data. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.910039

Ha, D. T., Loan, P. T. T., Giap, C. N., & Huong, N. T. L. (2020). An Empirical Study for Student Academic Performance Prediction Using Machine Learning Techniques. *International Journal of Computer Science and Information Security*, 18(3).

Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Huang, S., Brooks, M., Lee, M. J., & Asadi, H. (2019). Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. *AJR. American Journal of Roentgenology*, *212*(1), 38-43. https://doi.org/10.2214/AJR.18.20224

Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (2019). Student Academic Performance Prediction using Supervised Learning Techniques. *International Journal of Emerging Technologies in Learning (iJET)*, *14*(14), Article 14. https://doi.org/10.3991/ijet.v14i14.10310

Jagwani, A. (2019). A REVIEW OF MACHINE LEARNING IN EDUCATION. *JETIR*, *6*(5), Article 5. https://www.jetir.org/view?paper=JETIR1905658

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695. https://doi.org/10.1007/s12525-021-00475-2

Jawad, K., Shah, M. A., & Tahir, M. (2022). Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. *Sustainability*, *14*(22), Article 22. https://doi.org/10.3390/su142214795

Kaul, S., Fayaz, S. A., Zaman, M., & Butt, M. A. (2022). *Is Decision Tree Obsolete in Its Original Form? A Burning Debate*. 105-113. https://doi.org/10.18280/ria.360112

Lee, C. S., Cheang, P. Y. S., & Moslehpour, M. (2022). Predictive Analytics in Business Analytics: Decision Tree. *Advances in Decision Sciences*, *26*(1), 1-30.

Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2022). Predicting students' academic performance using machine learning techniques: A literature review. *International Journal of Business Intelligence and Data Mining*, *20*(4), 456-479. https://doi.org/10.1504/IJBIDM.2022.123214

Organisation for Economic Co-operation and Development. (2019). *PISA 2018 Results COMBINED EXECUTIVE SUMMARIES*. OECD Publishing.

Reddy, R. V. K., & Babu, D. U. R. (2018). A Review on Classification Techniques in Machine Learning. *International Journal of Advance Research in Science and Engineering, 7*(3).

Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, *2*, 100018. https://doi.org/10.1016/j.caeai.2021.100018

Rusdah, D. A., & Murfi, H. (2020). XGBoost in handling missing values for life insurance risk prediction. *SN Applied Sciences*, *2*(8), 1336. https://doi.org/10.1007/s42452-020-3128-y

Salas, F., & Caldas, J. (2024). Predicting undergraduate academic performance in a leading Peruvian university: A machine learning approach. *Educación*, 33(64), 55-85. https://doi.org/10.18800/educacion.202401.m003

Sandra, L., Lumbangaol, F., & Matsuo, T. (2021). Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review. *TEM Journal*, 1919-1927. https://doi.org/10.18421/TEM104-56

Senekane, M. (2019). Differentially Private Image Classification Using Support Vector Machine and Differential Privacy. *Machine Learning and Knowledge Extraction*, *1*(1), Article 1. https://doi.org/10.3390/make1010029

Sharma, N., Appukutti, S., Garg, U., Mukherjee, J., & Mishra, S. (2023). Analysis of Student's Academic Performance based on their Time Spent on Extra-Curricular Activities using Machine Learning Techniques. *International Journal of Modern Education and Computer Science*, *15*(1), 46-57. https://doi.org/10.5815/ijmecs.2023.01.04

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140-1144. https://doi.org/10.1126/science.aar6404

Singh, R., & Pal, S. (2020). Machine Learning Algorithms and Ensemble Technique to Improve Prediction of Students Performance. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(3), 3970-3976. https://doi.org/10.30534/ijatcse/2020/221932020

Soyoye, T. O., Chen, T., Hill, R., & Mccabe, K. (2023). Predicting Academic Performance of University Students Using Machine Learning: A Case Study in the UK. *2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 431-434. https://doi.org/10.1109/WI-IAT59888.2023.00071

Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). Behavioral Modeling for Mental Health using Machine Learning Algorithms. *Journal of Medical Systems*, *42*(5), 88. https://doi.org/10.1007/s10916-018-0934-5

Sudais, M., Safwan, M., Khalid, M. A., & Ahmed, S. (2022, enero 25). *Students' Academic Performance Prediction Model Using Machine Learning*. https://doi.org/10.21203/rs.3.rs-1296035/v1

Teoh, C.-W., Ho, S.-B., Dollmat, K. S., & Tan, C.-H. (2022). Ensemble-Learning Techniques for Predicting Student Performance on Video-Based Learning. *International Journal of Information and Education Technology*, *12*(8), 741-745. https://doi.org/10.18178/ijiet.2022.12.8.1679

Thanh Noi, P., & Kappas, M. (2017). *Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery*. http://resolver.sub.uni-goettingen.de/purl?gldocs-11858/7007

Turabieh, H., Azwari, S. A., Rokaya, M., Alosaimi, W., Alharbi, A., Alhakami, W., & Alnfiai, M. (2021). Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance. *Computing*, *103*(7), 1417-1438. https://doi.org/10.1007/s00607-020-00894-7

Vakili, M., Ghamsari, M., & Rezaei, M. (2020). *Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification* (arXiv:2001.09636). arXiv. https://doi.org/10.48550/arXiv.2001.09636

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11. https://doi.org/10.1186/s40561-022-00192-z

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zulfiker, M. S., Kabir, N., Biswas, A. A., Chakraborty, P., & Rahman, M. M. (2020). Predicting Students' Performance of the Private Universities of Bangladesh using Machine Learning Approaches. International Journal of Advanced Computer Science and Applications (IJACSA), 11(3), Article 3. https://doi.org/10.14569/IJACSA.2020.0110383